Chapter 6

Data

The E769 data set, comprised of approximately 400 million events written to tape at TPL during the 1987-88 Fermilab fixed target run, can be subdivided in a number of ways. First, the periodic structure of the secondary beams extracted from the primary 800 GeV p beam (i.e., spills of 22 second duration coming once every 60 seconds) provides the smallest unit into which events are grouped. Some variable number of spills (typically several hundred) are grouped into a larger unit called a "run", corresponding to a period of continuous data-taking under relatively uniform running conditions. About 2000 runs of data were taken during the experiment. Runs conducted under grossly similar running conditions are further grouped into "run regions", of which there are four (hereafter called "Regions 1-4").

During the first (second) half of E769, the mixed hadron beam was negatively (positively) charged. Regions 1 and 2 comprise the first half, 3 and 4 the second. During Region 1, the experiment ran at a beam energy of 210 GeV; in the subsequent three run regions, the beam was at 250 GeV. Within the positive running, the two run regions distinguish between the most characteristic DISC settings used during these periods. In Region 3, the DISC was typically set to identify kaons (also the case for the negative running); the DISC was set to identify protons in Region 4. Now that the run regions have been defined "historically", we will from now on refer to experimental conditions characterizing a particular spill, run, or run region in the present tense.

Once the data set has been broken down on the basis of running conditions, we further divide it into subsets on the basis of the identity of the beam particle[1] and the trigger type(s) under which the events were written to tape. These two criteria are interrelated through the inclusion of the DISC in the trigger logic. For the forward cross-section analysis, we require isolation of event subsets for each beam particle/beam energy combination present in the data. The trigger type(s) included in these subsets must be specified in order that absolute normalizations can be determined. These data subsets are defined in Section 6.2. For the differential cross-section analysis, the data sets used are largely the same, the main difference being that 210 and 250 GeV beam energy subsets are combined.

The analysis described in Chapter 5 is performed on these subsets for each D meson decay mode, resulting in the further subdivision of each subset into smaller event subsets, which in this case however are not disjoint. In some cases, subsets are then recombined when the resulting "composite" cross-sections retain some fundamental character (e.g., combining π^- and π^+ beam subsets). Signals for different D modes are also combined in the differential cross-section analysis. Motivations and justifications for such combinations of "initial-state" and "final-state" subsets are given in relevant sections of Chapters 8 and 9.

6.1 Event weeding

The data subsets used in both the forward and differential cross-section analyses undergo a "weeding" process whereby events are thrown out. In both analyses duplicate events are removed; this is "simple-weeding". For the forward cross-section calculations, absolutely normalized beam particle fluxes are obtained using scaler events. About 5% of events, however, are in spills for which the corresponding scaler event is missing, therefore rendering them unnormalizable. Another 3-5% of events are from spills or runs that are pathological in some way which makes their use suspect. For example, certain spills have incorrect NSPILL numbers which make them impossible

[1] As detailed in Section 3.2, a beam particle is considered positively identified if the probability that the tag is correct is at least 90%.

to associate with the correct scaler event. Some spills are thrown out for reasons other than pure bookkeeping; nonsensical average track multiplicities or DC efficiencies are examples of grounds for such disposal. Removal of events for the above reasons is called "full-weeding".

A small fraction of the time, scaler events indicate zero recorded flux for spills for which there are data events. Rather than throw out these events from anomalously-empty spills, flux correction factors χ_{empty} (see Section 8.1) are calculated by the following procedure: the ratio of the number of data events passing loose analysis cuts before and after removal of these "empty-spill" events is taken to be the ratio of the corresponding beam particle fluxes. The resulting factors are given in Table 6.1.

Data subset(s)	χ_{empty}
π^-,K^-; 210 GeV	0.982
π^-,K^-; 250 GeV	0.972
π^+,K^+	0.997
p	0.998

Table 6.1: "Empty-spill" flux correction factors.

6.2 Data subsets

The initial-state data subsets used in the forward cross-section analysis are tabulated in Table 6.2. For each beam energy/beam particle combination, the corresponding run region(s), trigger combination, and identified live beam particle flux (defined in Section 8.1) are given. Individual trigger types are defined in Section 3.5. Note that the proton beam subset consists of both the run region 3 and 4 components listed. Systematic errors associated with the flux totals are discussed in Section 8.2.3.

Beam particle	Beam energy (GeV)	Run region(s)	Trigger(s)	Identified live flux
π^-	210	1	ETπ+ETB	1.35×10^{10}
	250	2	ETπ+ETB	6.52×10^{10}
K^-	210	1	ETK	5.50×10^{8}
	250	2	ETK	1.67×10^{9}
π^+	250	3	ETπ+ETB+ETe	1.17×10^{11}
K^+	250	3	ETK	5.42×10^{9}
p	250	3	ETπ+ETB+ETe	5.99×10^{10}
		4	ETK	2.50×10^{9}

Table 6.2: E769 initial-state data subsets (absolute cross-section analysis).

6.3 Data signals

In general, the procedure for determining data signals is to perform (using FORTRAN-driven MINUIT) a log-likelihood fit, where the fit function is a bin-wise (10 MeV bin width) integration of a Gaussian over a background which is either flat or, in cases where a flat background does not provide a reasonable fit, linear. Masses are fixed at 94 PDG values. The width is fixed to the value returned by a fit of the total simply-weeded data signal, typically 1.5-2.5 MeV greater than that of the corresponding MC signal. These widths and masses are compiled in Table 6.3. In fitting relatively low-statistics signals (i.e., signals binned in x_F and p_T^2), the aforementioned procedure is modified, as described in Sections 6.3.1 and 6.3.2.

In Figs. 6.1 through 6.6, fitted invariant mass plots for each D mode are shown, including both the total signal (before any beam identification cuts are imposed) and its components broken down by beam particle; the resulting signal estimates, used in the forward cross-section calculations, are also given. Once events for which positive beam particle identification is impossible (~25%) are discarded, the data subsets for each D mode are made up of the following approximate beam particle fractions: 49% π^-, 17% π^+, 5% K^-, 13% K^+, and 16% p. The final event totals used in the forward

Decay mode	width (MeV)	mass (MeV)
$D^+ \to K\pi\pi$	11.0 ± 0.5	1869.4 ± 0.4
$D^0 \to K\pi$	12.9 ± 0.7	1864.6 ± 0.5
$D^* \to D^0\pi$	12.4 ± 1.1	2010.0 ± 0.5
$D_s \to KK\pi$	10.8 ± 1.1	1968.5 ± 0.7
Combined D	11.5 ± 0.4	—

Table 6.3: Data signal widths and (94 PDG) masses.

cross-section analysis are 994 ± 47 D^+, 847 ± 55 D^0, 100 ± 15 D_s, and 209 ± 19 D^*. As indicated above, the event totals used in the differential cross-section analysis are about 8% larger.

Note that in one case ($D^0 \to K\pi$, p beam), a linear background is not sufficient to give a reasonable fit to the invariant mass plot. Using a quadratic background goes too far the other way, resulting in a signal estimate which is unreasonably small. The compromise solution used is to fit the signal using a linear background over a more limited range (1.77-1.97 GeV); this fit (the one shown in Fig. 6.4) results in a signal estimate close to the average of those of the two aforementioned fits.

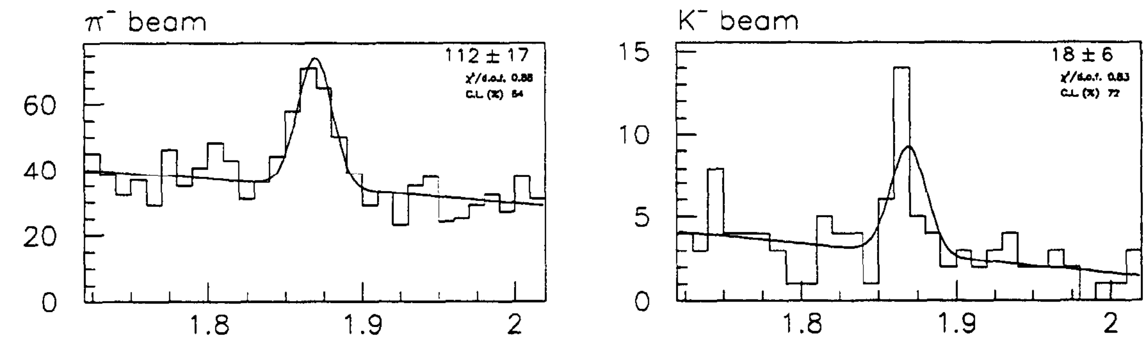

Figure 6.1: $D^+ \rightarrow K\pi\pi$ events vs. invariant mass (GeV), full-weeding, $x_F > 0$, 210 GeV beam.

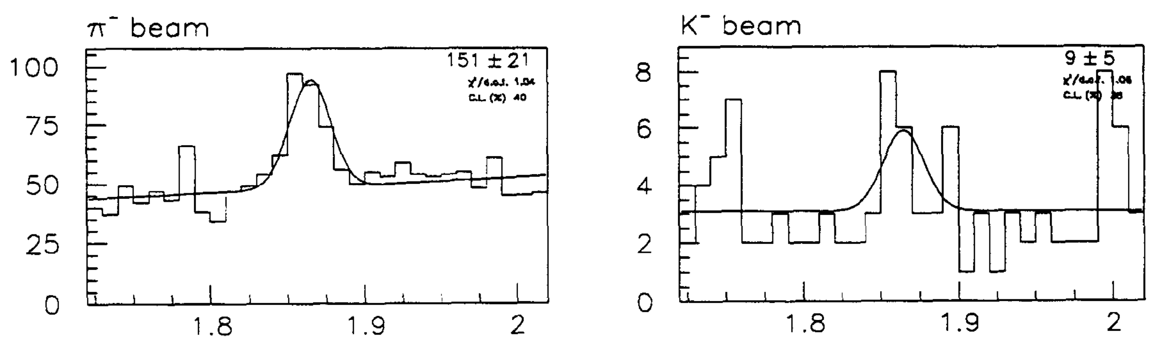

Figure 6.2: $D^0 \rightarrow K\pi$ events vs. invariant mass (GeV), full-weeding, $x_F > 0$, 210 GeV beam.

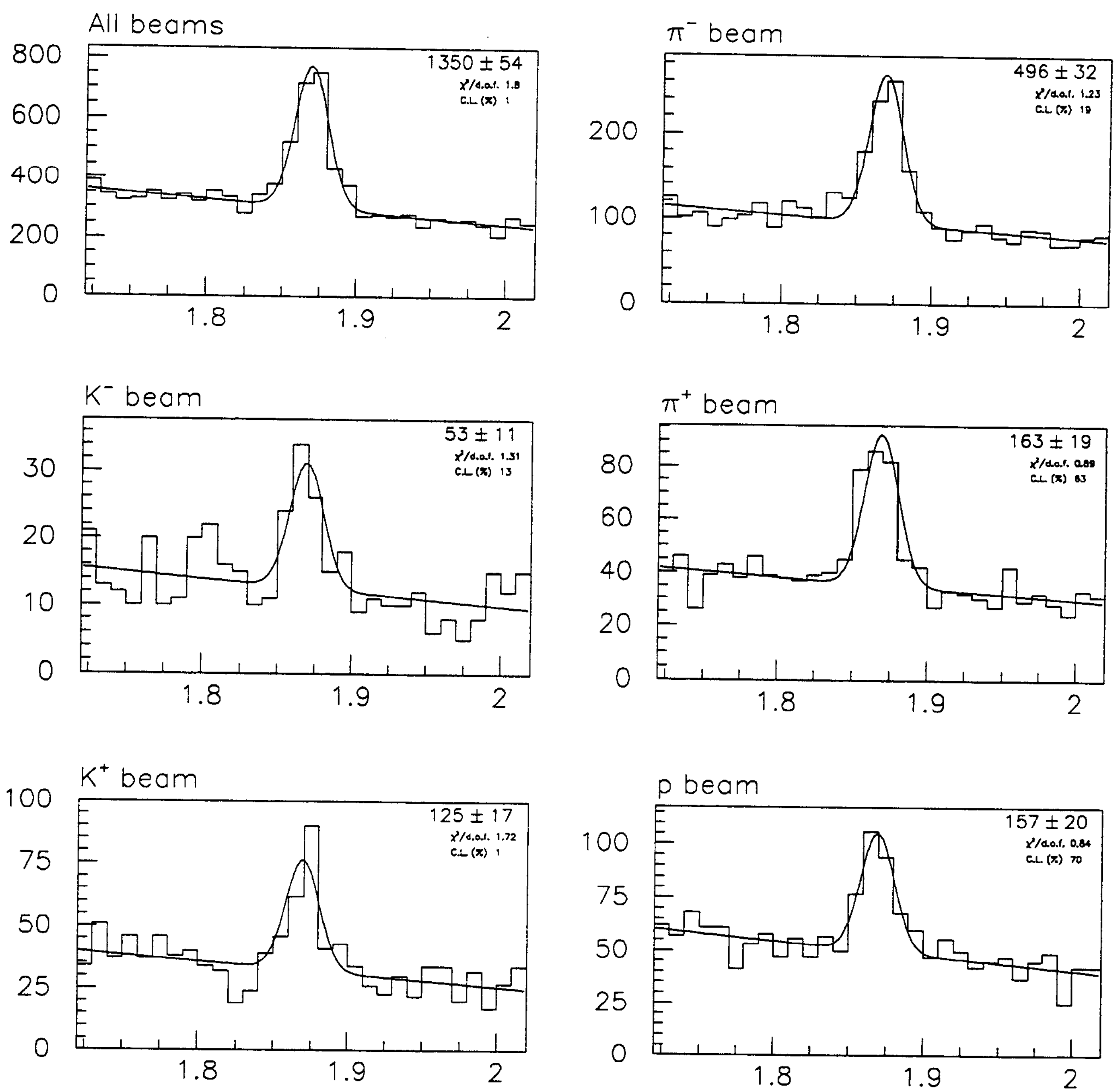

Figure 6.3: $D^+ \to K\pi\pi$ events vs. invariant mass (GeV), full-weeding, $x_F > 0$, 250 GeV beam.

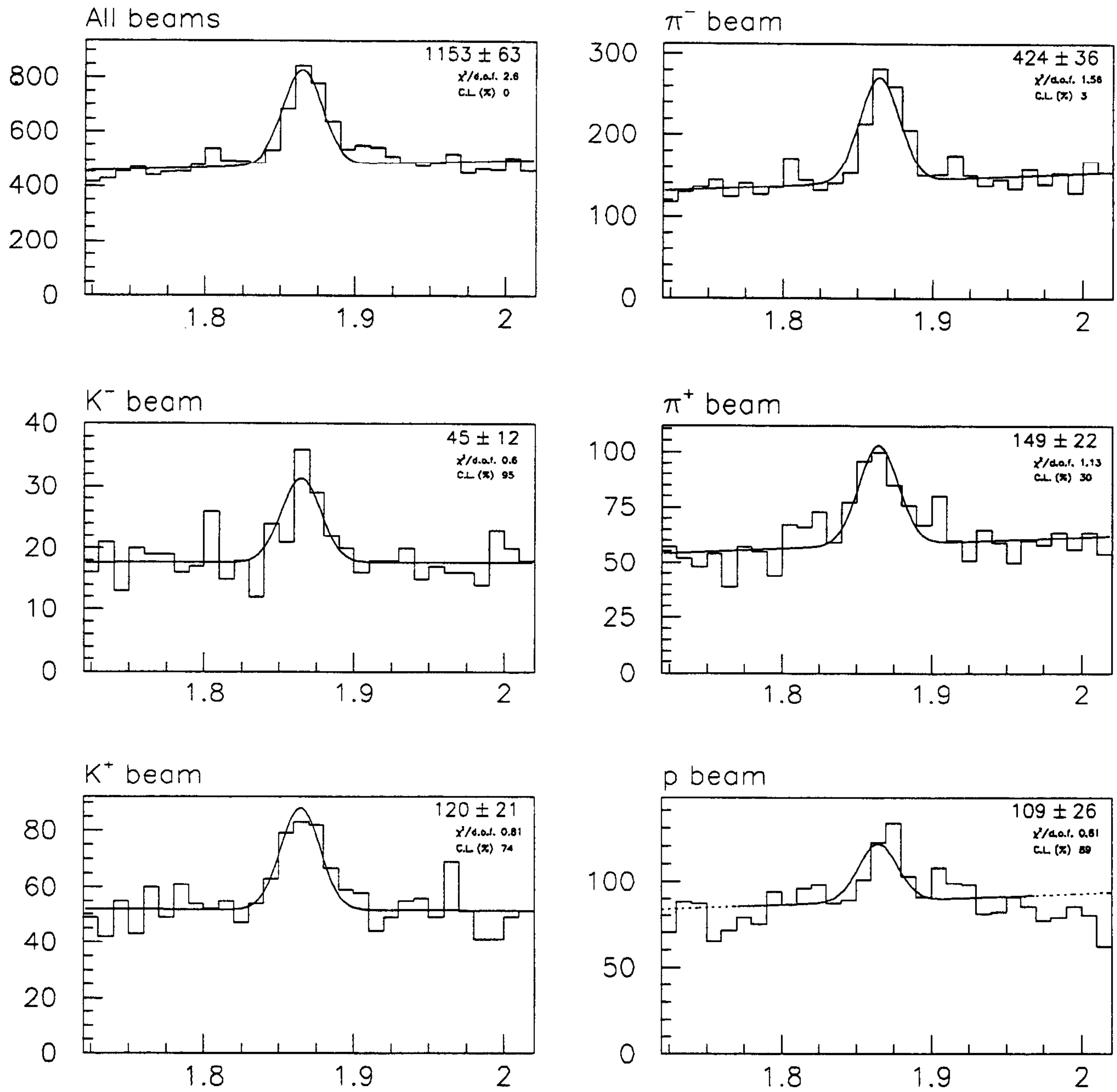

Figure 6.4: $D^0 \rightarrow K\pi$ events vs. invariant mass (GeV), full-weeding, $x_F > 0$, 250 GeV beam.

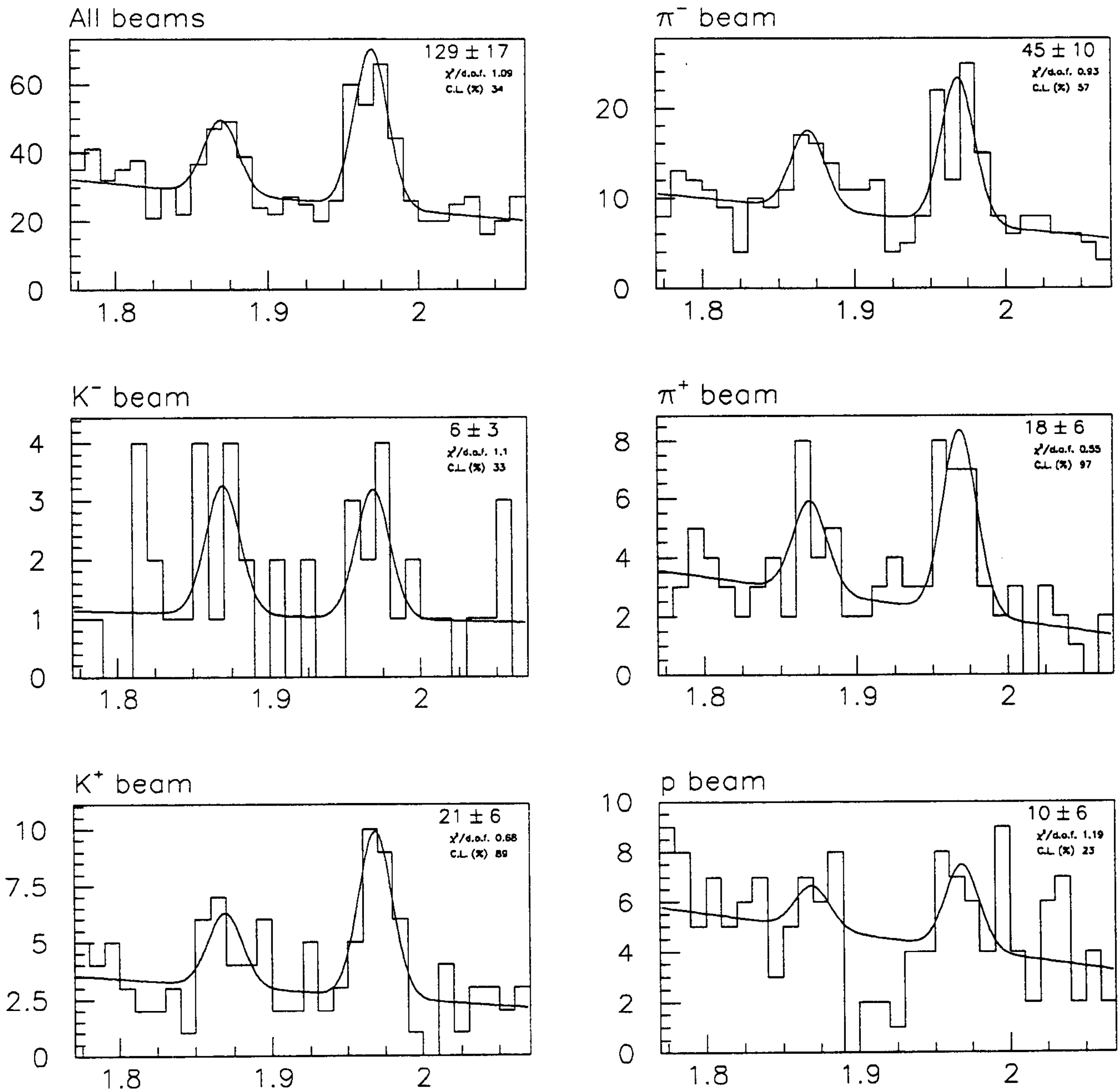

Figure 6.5: D^+ (left peak), D_s (right peak) $\rightarrow KK\pi$ events vs. invariant mass (GeV), full-weeding, $x_F > 0$, 250 GeV beam. The D^+ mode is Cabibbo suppressed.

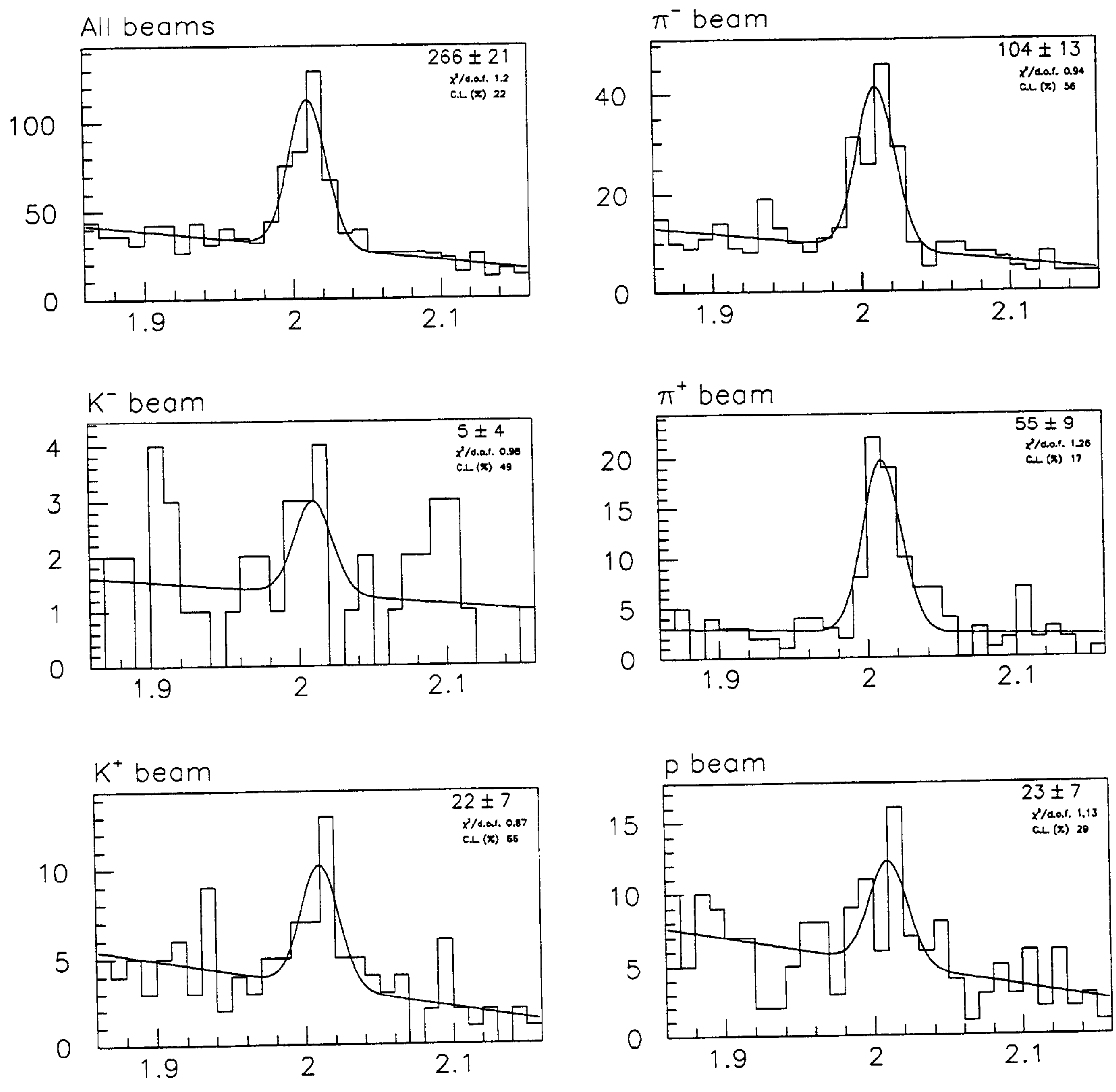

Figure 6.6: $D^* \to D^0\pi$ events vs. invariant mass (GeV), full-weeding, $x_F > 0$, 250 GeV beam.

6.3.1 Data signals vs. x_F

In fitting data signals, the x_F dependence of the signal widths is taken from the MC (see Section 7.3.1). An offset corresponding to the difference in the data and MC signal widths is added to the width versus x_F function, which is then used to obtain the values to which the widths are fixed during fitting. The constancy of signal mass versus x_F was checked in the D^+ and D^0 data; no indication of any x_F dependence is found (see Fig. 6.7).

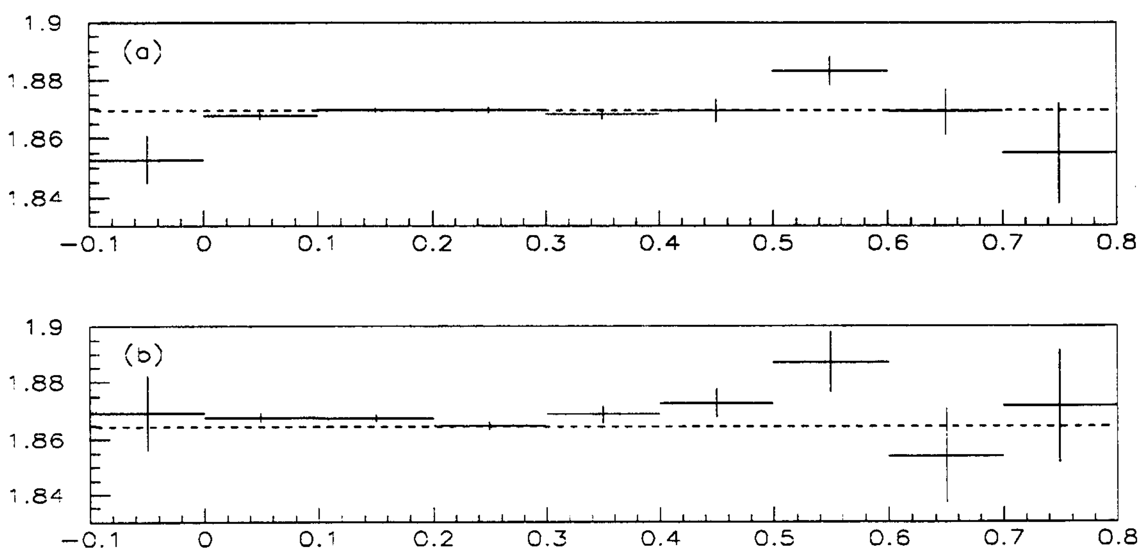

Figure 6.7: Data signal mass (GeV) vs. x_F for (a) $D^+ \rightarrow K\pi\pi$, π beam, and (b) $D^0 \rightarrow K\pi$, π beam. The dotted lines indicate the 94 PDG mass values.

As described in Chapter 9, differential cross-section results are obtained for a "combined D" sample comprised of the three pseudoscalar mesons D^+, D^0, and D_s; these species make up approximately 50%, 40-45%, and 5-10% of the combined D signals, respectively. Fits to the π^- beam combined D signals versus x_F, the highest-statistics representative of data signals thus broken down, are presented in Fig. 6.8.

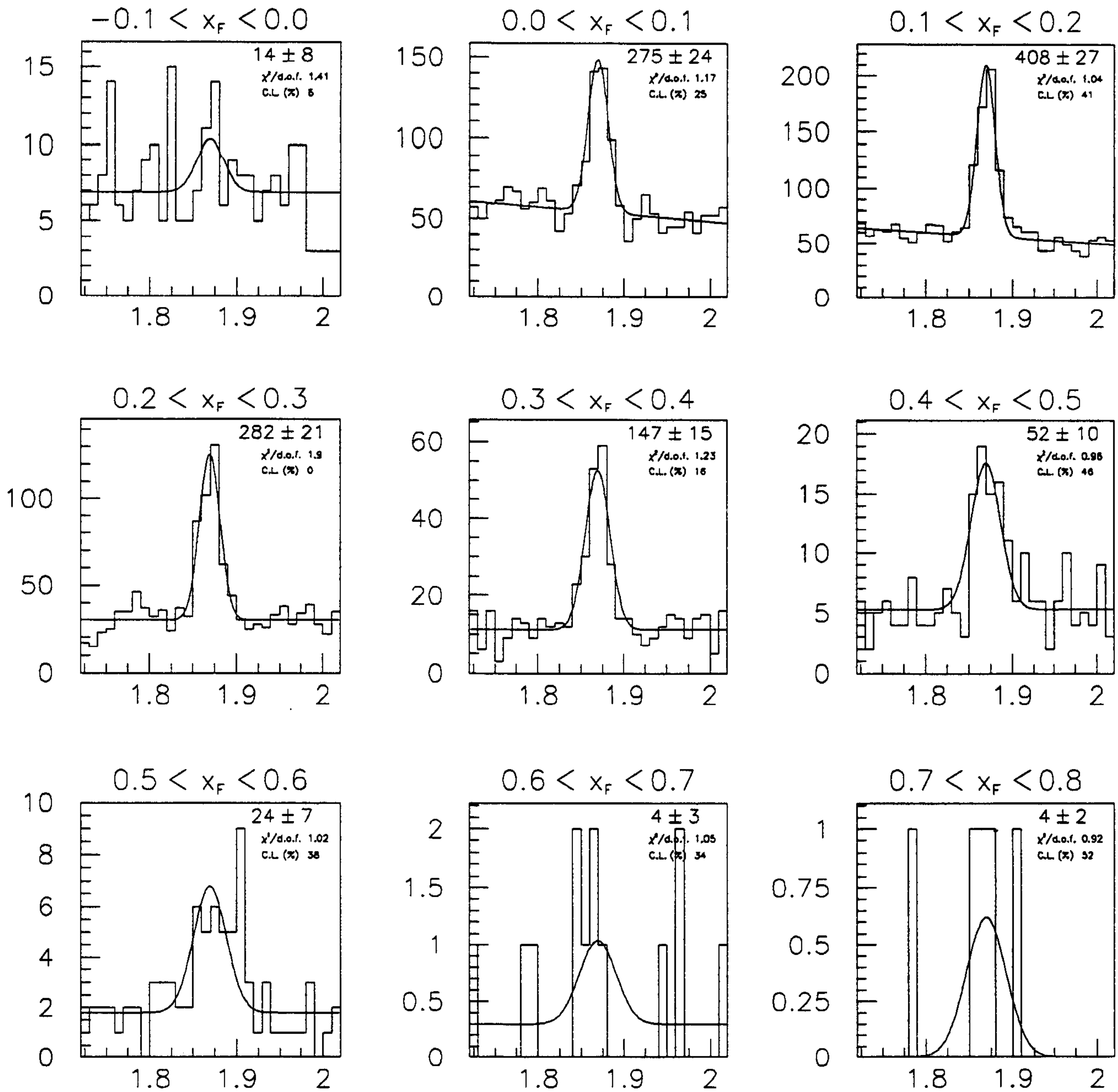

Figure 6.8: Combined D events vs. invariant mass (GeV), simple-weeding, π^- beam.

6.3.2 Data signals vs. p_T^2

In binning the data signals in p_T^2, it was found, owing to the exponential decrease in cross-section with rising p_T^2, that use of a single binwidth was incompatible with maintaining both sampling granularity and signal significance over the entire range for which data was available. Therefore, two binwidths, 1 and 2 GeV^2, are used for low and high p_T^2 respectively, the boundary between the two regions being determined on a case-by-case basis according to where the signals drop below the threshold of significance (2σ) used in fitting the differential distributions. This boundary ranges from 4 to 8 GeV^2. Once the binning for a particular mode/beam is chosen, it is used in all phases of the p_T^2 analysis. Fits to combined D signals versus p_T^2 are shown in Fig. 6.9.

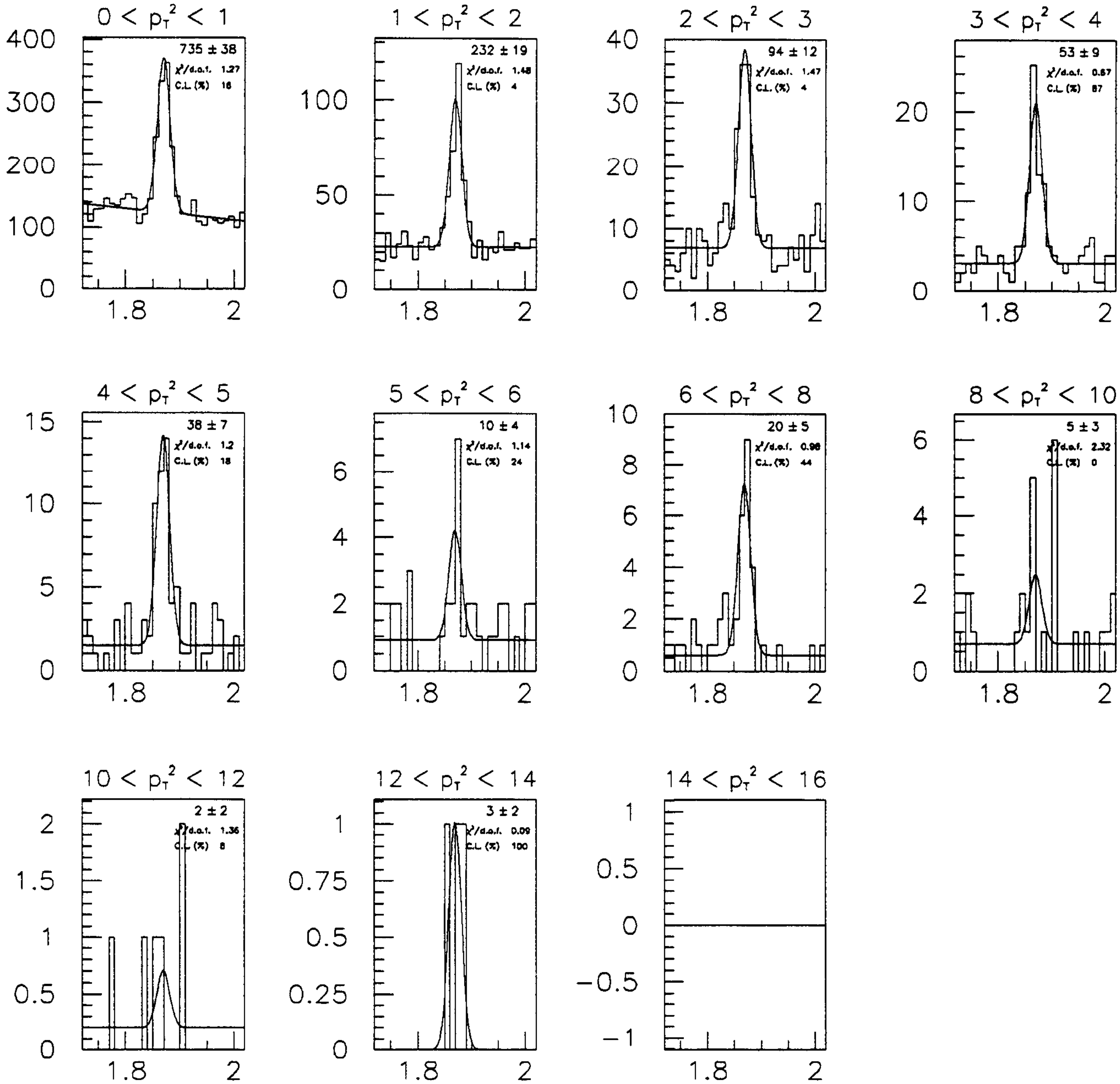

Figure 6.9: Combined D events vs. invariant mass (GeV), simple-weeding, $x_F > 0$, π^- beam.